Received March 22, 2019, accepted May 18, 2019, date of publication May 30, 2019, date of current version June 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920011

Ensembles of Patch-Based Classifiers for Diagnosis of Alzheimer Diseases

SAMSUDDIN AHMED[1], KYU YEONG CHOI[2], JANG JAE LEE[2], BYEONG C. KIM[2,3], GOO-RAK KWON[4], KUN HO LEE[2,5], AND HO YUB JUNG[1]

[1]Department of Computer Engineering, Chosun University, Gwangju 61452, South Korea
[2]National Research Center for Dementia, Chosun University, Gwangju 61452, South Korea
[3]Department of Neurology, Chonnam National University Medical School, Gwangju 61469, South Korea
[4]Department of Information and Communication Engineering, Chosun University, Gwangju 61452, South Korea
[5]Department of Biomedical Science, Chosun University, Gwangju 61452, South Korea

Corresponding author: Ho Yub Jung (hoyub@chosun.ac.kr)

This work was supported in part by the Brain Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning under Grant NRF-2014M3C7A1046050, and in part by the research fund from Chosun University, in 2018.

ABSTRACT There is ongoing research for the automatic diagnosis of Alzheimer's disease (AD) based on traditional machine learning techniques, and deep learning-based approaches are becoming a popular choice for AD diagnosis. The state-of-the-art techniques that consider multimodal diagnosis have been shown to have accuracy better than a manual diagnosis. However, collecting data from different modalities is time-consuming and expensive, and some modalities may have radioactive side effects. Our study is confined to structural magnetic resonance imaging (sMRI). The objectives of our attempt are as follows: 1) to increase the accuracy level that is comparable to the state-of-the-art methods; 2) to overcome the overfitting problem, and; 3) to analyze proven landmarks of the brain that provide discernible features for AD diagnosis. Here, we focused specifically on both the left and right hippocampus areas. To achieve the objectives, at first, we incorporate ensembles of simple convolutional neural networks (CNNs) as feature extractors and softmax cross-entropy as the classifier. Then, considering the scarcity of data, we deployed a patch-based approach. We have performed our experiment on the Gwangju Alzheimer's and Related Dementia (GARD) cohort dataset prepared by the National Research Center for Dementia (GARD), Gwangju, South Korea. We manually localized the left and right hippocampus and fed three view patches (TVPs) to the CNN after the preprocessing steps. We achieve 90.05% accuracy. We have compared our model with the state-of-the-art methods on the same dataset they have used and found our result comparable.

INDEX TERMS Alzheimer disease classification, ALZHEIMER disease detection, Alzheimer disease diagnosis, convolutional neural network, deep learning, machine learning, medical imaging.

I. INTRODUCTION

Alzheimer's disease (AD) is the most predominant neurodegenerative brain disease affecting elderly people worldwide and is considered to be one of the prime reasons for dementia [1]. AD is an irreversible, progressive neurobiological brain disorder and multifaceted disease of unknown etiology that slowly destroys brain cells, causes memory and thinking skill losses, and ultimately accelerates the loss of ability to carry out even the simplest tasks [2]. The cognitive decline caused by this disorder progresses toward dementia. The disease progresses over time from its initial stage of normal controlled (NC) to mild cognitive impairment (MCI) and then eventually reaches the AD affected stage [3].

According to the report provided by the Alzheimer's Association [1], 60 million people are predicted to be affected by AD within the next 50 years. The estimation provided by the World Alzheimer Report [4] is more alarming. One person is becoming affected by dementia every three seconds, 60% of whom are affected because of AD. The total estimated patient growth is 152 million by 2050, costing 2 trillion per annum by 2030 [4].

Studies are ongoing for the early diagnosis of this disease in order to put a brake on the abnormal degeneration of the brain, to reduce the cost of patient care and to ensure better management. Previous studies [5]–[8] have shown

The associate editor coordinating the review of this manuscript and approving it for publication was Carlo Cattani.

that machine learning algorithms were able to classify AD more accurately than experienced clinicians [3]. Recently, deep learning has shown outstanding performance in classification and regression [9]. Therefore, deep learning-based approaches are becoming the obvious choice for the detection of Alzheimer diseases.

The state-of-the-art approaches either consider the whole brain in a single modality [10], [11] or multimodal [6] datasets to train machine learning models, which have been shown to demonstrate greater accuracy than manual diagnosis. Magnetic resonance imaging (MRI), positron emission tomography (PET), biospecimens, genotyping and sequencing data and clinical datasets are different modalities for AD research. Investigating more than one data modality is time consuming and expensive. Moreover, modalities such as PET may have radioactive side effects on patients. Here, we consider structural magnetic resonance imaging (sMRI) as the modality of our experiments for the following advantages: 1) high degree of imaging flexibility; 2) high tissue contrast; 3) no need for ionizing radiation; 4) useful information about the anatomy of the brain;

The accuracy of an AD diagnosis mostly depends on the biomarkers of the disease. A biomarker serves as a determinant of health and disease; it is measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention [12]. Studies have revealed that there is an association between structural changes of the AD brain and cognitive loss. Hippocampal shrinkage is observed in the preliminary stages of AD, and it has a proven correlation with memory impairment. As reported in [13], the yearly brain atrophy rate for AD patients is 2.4% ± 1.1%, whereas the figure is 0.5% ± 0.4% for age-matched control subjects with normal brains.

The hippocampal atrophy rate is higher than that of the whole brain. According to [14], the average hippocampal atrophy rate for AD patients was approximately 4 to 6%, and the rate was 1 to 2% for age-matched control subjects with normal brains. Thus, we focus on the hippocampus region as the input feature for the convolutional neural network (CNN). We localize the hippocampus manually for each MRI; then, we generate 32 × 32 patches from the localized region.

Our prime objective of this research is to design a simpler CNN. We have avoided feeding the entire MRI volume into the network to reduce unnecessary computations. The approach improved the computational efficiency as well as the prediction accuracy. It has also resolved problems due to data scarcity that are associated with deploying CNNs. To achieve our objective, we have generated 32 × 32 patches from each of the sagittal, axial and coronal views and merged them as a single sample. These three view patches (TVPs) are fed into the CNN.

In this paper, we have summarized related works in section II. Data collection and preparation are outlined in section III. Our framework is described in section IV. The experimental setup is illustrated in section V. The results are discussed in section VI, and section VII concludes the work.

II. RELATED WORKS

There are studies based on the conventional machine learning techniques which focused on developing models to detect anatomical and functional disorders due to AD in human brain [15]–[21]. These methods have primarily relied on manually designed features, which heavily depend on professional expertise, require repeated trials, and tend to be time-consuming and subjective processes. However, as the cause of AD is not completely understood, designing robust analysis methods for effective hand-crafted features using medical experts' knowledge is a challenging task. In contrast, deep learning is capable of automatically learning input features from a large set of training data. Many previous studies were conducted to further explore CNN architectures dedicated to generating robust AD features.

Gupta *et al.* [8] used cross-domain features to represent MRI data. They deployed a stacked autoencoder (SAE) to learn a set of bases from natural images and then applied a CNN to obtain a more effective feature representation for AD classification. Despite being very simple, they showed high classification performance in comparison with contemporary approaches. Liu *et al.* [2] also proposed an SAE-based multimodal neuroimaging feature learning algorithm from a region of interest (ROI) for AD diagnosis. This framework uses a zero-masking strategy for data fusion to extract complementary information from multiple data modalities.

Brosch and Tam [22] learned a low-dimensional manifold of brain volumes with a deep belief networks (DBN) algorithm to detect the modes of variations that correlate to demographic and disease parameters for AD. Their primary contributions are following: 1) they introduced a much more computationally efficient training method for DBNs that allows training on 3D medical images with a resolution up to 128 × 128 × 128, and 2) they demonstrated that DBNs can learn a low-dimensional manifold of brain volumes that can detect modes of variations.

Payan and Montana [23] used a sparse autoencoder to learn feature embedding and then feed these embeddings to a convolution neural network for AD classification. The authors built a learning algorithm that is able to discriminate between healthy brains and diseased brains using MRI images as input. They investigated a class of deep artificial neural networks and a specific combination of sparse autoencoders and CNNs. The main novelty of their approach is to use 3D convolutions on the whole MRI image. Li *et al.* [7] proposed a robust multitask deep learning framework using a dropout [24] and stability selection technique to improve the ROI feature representation for AD/MCI diagnosis.

Shi *et al.* [6] developed a robust deep learning framework for multimodal AD diagnosis from MRI and PET scans. They applied principal component analysis (PCA) to obtain features and then utilized a stability selection technique together with the least absolute shrinkage and selection operator (LASSO) method [5] to select the most effective features. The selected features were then processed by the deep learning structure. Model weights in the deep structure were first

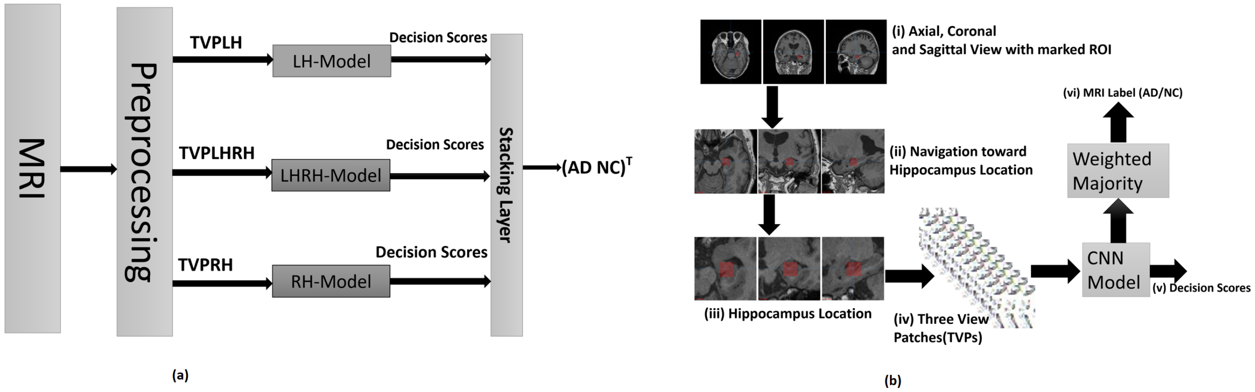

FIGURE 1. Pipeline for AD/NC classification from structural magnetic resonance imaging (sMRI). (a) Outline of the overall pipeline of the proposed method: Preprocessing includes intensity normalization, hippocampus localization and TVP generation. The size of the TVP from the left hippocampus (TVPLH) and right hippocampus (TVPRH) is 32 × 32 × 3; the size of merged TVP from both the hippocampi (TVPLHRH) is 64 × 32 × 3; left hippocampus classifier (LH-Model) and right hippocampus classifier (RH-Model) are pretrained CNN models that take TVPs as input and yields a softmax score for each TVP; both hippocampi classifier(LHRH-Model) is another pretrained CNN model that takes TVPLHRH of size 64 × 32 × 3 and yields a softmax score; The scores are summed up and normalized by a softmax classifier in stacking layer to obtain the final label; (b) demonstrates the workflow of an individual model: (i) three view planes of an MRI drawn from the center of the MRI; (ii) global view of hippocampus in three different views; (iii) closer views of hippocampus locations in three different views; (iv) TVPs of size 32 × 32 × 3; (v) decision scores yielded by patch-based classifiers are fed to the stacking layer for ensemble classification; and (vi) MRI label yielded by individual models based on collections of TVPs.

initialized by unsupervised training and then fine-tuned by AD patient labels. During the fine-tuning phase, the dropout layer was deployed to improve the model's generalization capability. Finally, the learned feature representation was used for AD/MCI classification by a support vector machine (SVM).

III. DATA COLLECTION AND PREPROCESSING

In this section, we describe the dataset that we have used in our study. The preprocessing steps performed on the data also discussed here.

A. DATASET

Magnetic resonance imaging (MRI) is the de facto modality in brain studies due to its superior image contrast in soft tissue without involving ionizing radiation. MR images are widely used to examine other anatomical regions as well [25]. There are many MRI datasets for AD studies, such as ADNI, BgBrain, OASIS, and AIBL.

Data used in the preparation of this article were obtained from the ADNI [1] database (adni.loni.usc.edu). ADNI was launched in 2003 as a public-private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. For up-to-date information, see www.adni-info.org.

We have used 352 MRI scans of the ADNI dataset that are separated into three different classes: AD, NC and MCI. Among these classes, we used only AD and NC data. There are 77 MRI scans labeled as AD, 129 are labeled as NC, and the remaining 146 are labeled as MCI. The size of the MRI volume is 170 × 256 × 256 in most of the cases. However, the large 3D size does not hinder the training as we consider only hippocampus regions.

After training, validation and testing with ADNI data, we have retrained and tested our models with Gwangju Alzheimer's and Related Dementia (GARD) cohort dataset provided by the National Research Center for Dementia (NRCD), Gwangju, Republic of South Korea. The GARD dataset has 326 baseline MRI scans. All scans are from Korean patients with an age range of 49 years to 87 years. There are four labels in the dataset, namely, AD, NC, mAD and aAD. AD, NC and mAD are similar to the ADNI database classes of AD, NC and MCI, respectively. aAD is another stage known as asymptomatic AD, where there are no symptoms of AD or MCI in terms of recognition capability but the patients are biomarker positive [26], [27].

In the GARD database, there are 81 samples in the AD class, whereas the number is 171 for the NC class. The number of samples for the mAD and aAD classes is 39 and 35, respectively. We followed the same data separation rule and preprocessing techniques for the ADNI and GARD datasets.

B. MRI PREPROCESSING

The ADNI dataset was already corrected for intensity inhomogeneity. The axial, coronal and sagittal views of a sample MRI are shown in Figure 1(b). We have normalized the intensity values by subtracting the mean intensity and then dividing by the standard deviation to have zero mean and unit

[1] The Alzheimer's Disease Neuroimaging Initiative data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (*http://adni.loni.usc.edu/*). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

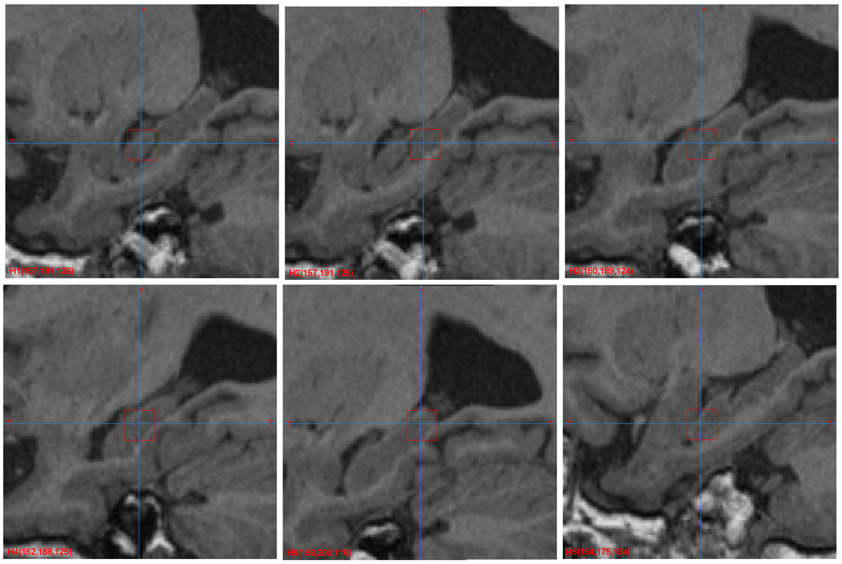

FIGURE 2. An example of selected six points (viewed on the sagittal plane) that are used as the center of 8 × 8 × 8 cubes for generating reference points. The reference points are selected from the cube by a semirandom process. TVPs are generated on the reference points.

variance of the input. The normalization is defined in (1).

$$\widetilde{I}(i_x, i_y, i_z) = \frac{I(i_x, i_y, i_z) - \mu(I)}{\sigma(I)} \tag{1}$$

Here, $I(i_x, i_y, i_z)$ is the intensity of (i_x, i_y, i_z) location before normalization, $\mu(I)$ is the mean intensity and $\sigma(I)$ is the standard deviation of the intensity; $\widetilde{I}$ is the normalized intensity of the MRI.

After intensity normalization, we manually observed hippocampus locations (h_x, h_y, h_z) on the normalized MRI by using Mango, a multi-image analysis graphical user interface (GUI) [28]. Six example locations are presented in fig. 2. Each location and its neighboring points up to α, β and γ pixels in sagittal, coronal and axial direction, respectively, were used as a reference frame for patch generation. Careful selection of these shape constants; i.e., α, β, and γ; ensured that each reference frame lies within the hippocampus region. In our experiment, we have selected $\alpha = \beta = \gamma = 4$. Different values of these constants provides flexible shape of the reference frame to adjust with the shape of ROI. We have randomly chosen n_x, n_y, and n_z number of co-ordinates in sagittal, coronal, and axial directions, respectively. As described in (2), (3), (4) and (5), these co-ordinates are used to generate reference points for TVPs.

$$rClT_x = rand(h_x - \alpha, h_x + \alpha, n_x) \tag{2}$$

$$T_y = rand(h_y - \beta, h_y + \beta, n_y) \tag{3}$$

$$T_z = rand(h_z - \gamma, h_z + \gamma, n_z) \tag{4}$$

Here, T_x, T_y, T_z represent uniformly distributed integer samples from the specified interval. The number of samples drawn from the interval are denoted by n_x, n_y, and n_z respectively. $rand(h_x - \alpha, h_x + \alpha, n_x)$ returns n_x number of uniformly distributed random integers from the interval $(h_x - \alpha, h_x + \alpha)$. The same explanation follows for (3) and (4).

The reference points were generated by taking all (i, j, k) tuples of $T_x \times T_y \times T_z$ (i.e., the cartesian product). The disjunction of all reference points obtained from each reference frame were used to generate TVPs. The equation (5) summarized the operation. The algorithm 1 concisely describes the reference point generation process.

$$R = R \cup \{\forall (i, j, k) | i \in T_x, j \in T_y, k \in T_z\} \tag{5}$$

To avoid data imbalance between two classes, we took relatively fewer samples from the MRIs of the NC class. Sample reduction for NC class was done by reducing the values of shape constants of reference frame in (2), (3), and (4). As our patch size is 32×32 smaller number of samples did not affects in exploring the whole hippocampus volume.

There are three different patches in each sample for individual hippocampus classification. Patches were taken from each of the three orthogonal axial, coronal, and sagittal view planes for each reference point. For combined classification of the left and right hippocampus, the sample consists of six patches.

IV. METHODOLOGY

The proposed pipeline is depicted in Figure 1(a). Our framework consists of three individual models for generating decision scores on individual patches, followed by a score aggregator and final classifier.

After collecting data, we performed the preprocessing tasks, as stated in the previous section. Then, we performed manual localization of the left and right hippocampus, which we consider as the ROI for our experiment. Then, from the ROI, we generated TVPs of size 32 × 32 × 3 or 64 × 32 × 3.

Algorithm 1 Algorithm for Reference Point Generation

```
Input: H = {H1, H2, H3, H4, H5, H6}: Manually
       observed approximately equidistant locations
       inside the hippocampus
Output: R: a set of reference locations,(x,y,z)
$R = \{\}$
for each point $H_r(h_x, h_y, h_z) \in H$ do
    $T_x = rand(h_x - \alpha, h_x + \alpha, n_x)$
        // $rand(h_x - \alpha, h_x + \alpha, n_x)$ returns $n_x$ number of
    uniformly distributed random integers from
    the interval $(h_x - \alpha, h_x + \alpha)$
    $T_y = rand(h_y - \beta, h_y + \beta, n_y)$
    $T_z = rand(h_z - \gamma, h_z + \gamma, n_z)$
    $R = R \cup \{\forall(i, j, k) | i \in T_x, j \in T_y, k \in T_z\}$
return R
```

We trained three individual models with these generated patches. At first, we designed the classifier for left hippocampus. Then, we tried the architecture for right hippocampus classification. But, after several trial, we found the right hippocampus classifier with even simpler network. As the input size is different for both hippocampi classifier(i.e., LHRH model), we had to tweak the architecture of the related model. The performance of each model was measured individually. These three models were then added together, and a softmax classifier was used for the final distinction.

A. PATCH-BASED CLASSIFIER

It is well known that CNNs are highly susceptible to the sample size. The more the samples we have from each class, the more accurate the CNN performs. Classification accuracy is subject to the discriminating features among the available classes [29]. The availability of discriminating features of a class depends on the number of samples from the class. The main problem of AD diagnosis is the scarcity of data. We have a limited number of samples from each class. This scarcity of data may lead to an overfitted model. Therefore, we deployed a patch-based classifier because we can generate a sufficient number of patches for training.

In our proposed patch-based classification architecture, shown in Fig 1(a), the input to the CNN is TVP. Each TVP is centered at the locations generated from the reference points. The CNNs are trained to predict individual TVPs as AD or NC. Based on the collections of individual TVP decisions, an MRI is classified into two objective classes. We keep kernel sizes less than or equal to 5 × 5 to extract detail information over the hippocampus. We use the rectified linear unit (ReLU) [30] as the activation function. The pooling operation [31] selects the activation from rectangular areas of specified size; it downsamples the patches by a factor of defined stride size. Max pooling and/or average pooling were used in the networks. We use batch normalization [32] before each convolution layer to enforce a normal distribution at the output of the layer.

1) CNN FOR THE LEFT HIPPOCAMPUS

The model for the left hippocampus classification is presented in Table 1. The output of the third convolution layer is the feature embedding of the left hippocampus region. These features are further fed to the fully connected layers to classify AD versus NC. Adding a dropout of 0.75 in the first fully connected layer slightly improved the accuracy. We used softmax as the last layer activation and cross-entropy as the loss function. The Adam optimizer [33] and Xavier initialization [34] were used. The exponential decay rate for first and second moment estimates are 0.9 and 0.999 respectively. The architecture and structural details of the proposed CNN are noted in Table 1. Total number of trainable parameter in the network is 105,826.

2) CNN FOR THE RIGHT HIPPOCAMPUS

The CNN architecture for the right hippocampus classification is illustrated in Table 1. We tried different structures and hyperparameters. We determined the proposed network after several trials. There are three convolution layers and two fully connected layers in the model. Each convolution layer and fully connected layer are preceded by batch normalization and followed by the average pooling layer. The first convolution layer does not have the batch normalization, but inputs are normalized previously. Before the last fully connected layer, we used a dropout of 0.25, which converges the training process faster and increases the accuracy. The output of the last convolution layer is the feature embedding of the right hippocampus region. The optimization and weight initialization are the same as the model used for the left hippocampus classification. The architecture and structural details of the proposed CNN for the right hippocampus classification are noted in Table 1. Total number of parameters in the network is 100,197 among which 99,925 parameters are trainable.

3) CNN FOR THE LEFT AND RIGHT HIPPOCAMPUS CLASSIFICATION

The architecture of the proposed CNN for the classification of both hippocampi is shown in Table 1. There are seven convolution layers. Each follows batch normalization and/or drop out. It takes input of size 64 × 32 × 3. We merged the TVPs of size 32 × 32 × 3 from the left and right hippocampus to generate these input patches. We illustrate the merging operation in Figure 3. The output of the seventh convolution layer is the feature embedding of the hippocampus region. These features are further fed to the fully connected layers to classify AD versus NC. The total number of parameters for this network is 409,666. The Adam optimization and Xavier initialization techniques were used in this model.

B. ENSEMBLE CLASSIFICATION

The reasoning behind using an ensemble is to bypass the weakness of individual models. Each model has its own hypothesis about the given input. By stacking different models with different assumptions about a class label, we can find

TABLE 1. CNN architecture for the classification of left hippocampus(LH Model), right hippocampus(RH Model) and both hippocampi(LHRH Model).

RH Model		LH Model		LHRH Model	
Layer(Type)	**Parameters**	**Layer(Type)**	**Parameters**	**Layer(Type)**	**Parameters**
Conv.	6×5×5×3+6	Conv.	64×5×5×3+64	Conv.	32×3×3×3+32
Avg. Pool.	2×2	Max. Pool.	5×5	Batch Norm.	128(64)
Batch Norm.	24(12)	Conv.	32×5×5×64+32	Conv.	32×5×5×32+32
Conv.	16×5×5×6+16	Max. Pool.	5×5	Batch Norm.	128(64)
Avg. Pool.	2×2	Conv.	16×5×5×32+16	Conv.	32×5×5×32+32
Batch Norm.	64(32)	Max. Pool.	5×5	Batch Norm.	128(64)
Conv.	120×5×5×16+120	Flatten	-	Dropout	0.4
Batch Norm.	480(240)	Fully	16×2304+16	Conv.	64×3×3×32+64
Flatten	-	Dropout	0.75	Batch Norm.	256(128)
Fully	9720×5+5	Fully	2×16+2	Conv.	64×5×5×64+64
Dropout	0.25	Output	$[AD\ NC]^T$	Batch Norm.	256(128)
Fully	2×5+2			Conv.	64×5×5×64+64
Output	$[AD\ NC]^T$			Batch Norm.	256(128)
Training details of all models: Weight Initialization: Xavier; Optimization: Adam; Exponential decay rates,(β_1,β_2), for LH, RH, and LHRH models are: (0.99,0.999),(0.98,0.9998) and (0.99,0.999), respectively; Divide by zero prevention constant: 0.0001; Batch Size: 32; Total trainable parameters for the models are: (RH; LH; LHRH)= (99,925; 105,826; 409,986).				Dropout	0.4
				Conv.	128×4×1×4×64+128
				Batch Norm.	512(256)
				Flatten	-
				Dropout	0.4
				Fully	2×1152+2
				Output	$[AD\ NC]^T$

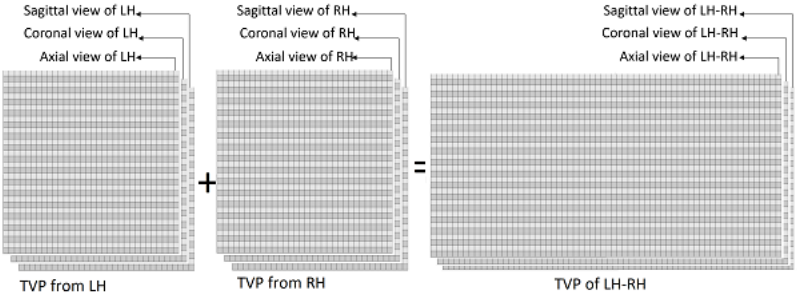

FIGURE 3. To prepare training and testing samples (of size 64 × 32 × 3) for the patch-based left and right hippocampus classifier (LHRH model), each TVP (of size 32 × 32 × 3) from the left hippocampus is merged with the corresponding TVP (of size 32 × 32 × 3) of the right hippocampus.

a better classification that may not be possible with individual models.

To determine the most appropriate class label for a given patch, the results from all three models are combined. Each patch-based model produces decision scores of a patch that indicate how well the patch fits a class. The individual decisions of the relevant patches are combined in the stacking layer. Then, a softmax is applied to find normalized scores. The most likely value is selected as the final class for a patch. Two reference points are considered for a single score from the ensemble layer: one from the left hippocampus and one from the right hippocampus. The patches are classified by related patch-based classifiers. The final scores of patches are forwarded to stacking layer.

From each test MRI, at least 32 pairs of TVPs are generated for classification. The majority decision from the stacking layer determines the class label of each MRI. This classification is grounded on weighted majority voting. Adding all the scores from each classifier ensures robust accuracy. The classification process is described in algorithm 2

V. EXPERIMENTAL SETUP

A. DATASET SEPARATION

From the ADNI dataset, we consider only those subjects whose disease status remains the same over different MRI scans. We have selected a total of 60 subjects from our dataset. For each subject, there are different MRI scans. We separate the training, testing and validation set in such a way that the conjunction of any two sets, keeping the subject ID of MRI scans as the key, yields the null set. This ensures the prevention of data leakage. We also ensure that MRI scans from each class are uniformly distributed among the three

Algorithm 2 Algorithm for Ensemble Decisions

```
Input: Data: MRI Image Volume; $R_L$, $R_R$: two sets of
       reference locations,(x,y,z) for left and right
       hippocampus, respectively.
Output: $DS(DS[ad], DS[nc])$: Decision scores of an
        MRI
Data: Let $N = |R_L| = |R_R|$ be the number of patches
      sampled from a hippocampus of an MRI.
      LHMODEL(), RHMODEL(), LHRHMODEL()
      returns the decision scores for individual TVPs as
      $(s[ad], s[nc])$
$tvplh = TVP\_Generator(R_L)$
    // $TVP\_Generator(R_L)$ returns $32 \times 32 \times 3$ TVP
    centering at the locations $\in R_L$
$tvprh = TVP\_Generator(R_R)$
$tvplhrh = merged\_TVP\_Generator(R_L, R_R)$
    // $merged\_TVP\_Generator(R_L, R_R)$ returns TVPs of
    size $64 \times 32 \times 3$ generated from pairs of TVPs of
    size $32 \times 32 \times 3$ centering at the pair of
    locations (l,r). Here $l \in R_L$, $r \in R_R$ and the
    function for mapping the corresponding
    locations is, $F : R_L \rightarrow R_R$ is one to one and onto.
$s_l = LHMODEL(tvplh)$
$s_r = RHMODEL(tvprh)$
$s_{lr} = LHRHMODEL(tvplhrh)$
$score_{[ad]} = \sum_{m \in \{l,r,lr\}} \sum_{i=1}^{|N|} \mathbf{s}_m^{(i)}[ad]$
$score_{[nc]} = \sum_{m \in \{l,r,lr\}} \sum_{i=1}^{|N|} \mathbf{s}_m^{(i)}[nc]$
$DS[ad] = \frac{e^{score_{[ad]}}}{e^{score_{[ad]}} + e^{score_{[nc]}}}$
$DS[nc] = \frac{e^{score_{[nc]}}}{e^{score_{[ad]}} + e^{score_{[nc]}}}$
return DS
```

sets to address the class imbalance problem. We keep 60% of MRI scans as the training set, 20% for the test set and 20% for the validation set. We augmented the data of each class by applying shearing, rescaling and zooming of the patches.

All the MRIs in the GARD dataset are baseline MRI scans, so we did not need to separate the MRIs according to patients. We divided the dataset into training, validation and a test set according to the procedure that we followed for the ADNI dataset separation. We also applied the same data augmentation techniques to the GARD dataset.

B. PLATFORM

We use the TensorFlow GPU 1.8, keeping Keras as the backend, on top of the Python 3.6 environment. An Intel(R) Xeon (R) CPU E5-1607 v4 @ 3.10 GHz with a 32 GB RAM machine was used. The GPU was NVIDIA Quadro M4000.

C. TRAINING

For patch-based classification, we trained different architectures with different hyperparameters. The presented models were trained for 20 epochs with a batch size of 32. We followed the 60-20-20 approach for using sample patches for training, validation and testing. We started the training with a learning rate of 0.001. If the validation loss stopped improving for 3 consecutive epochs, we reduced the learning rate by a factor of 10. It was observed that the learning rates were between 0.001 to 0.0001. The default parameter settings were used for the optimizers, regularizers and constraints.

VI. RESULTS

For evaluating the models, we have taken $accuracy = \frac{TP+TN}{TP+TN+FP+FN}$, $precision = \frac{TP}{TP+FP}$, $recall = \frac{TP}{TP+FN}$ and $f1\ score = 2 \times \frac{precision \times recall}{precision + recall}$ into consideration. Here, TP, TN, FP and FN are acronyms for the number of model-predicted true positive, true negative, false positive and false negative samples, respectively. For evaluating each model, we used an individual MRI as a sample. We generated at least 32 TVPs for each test MRI to obtain its label. First, we feed TVPs to the patch-based classifier to obtain the decision scores for each individual TVP. We then added the scores of all patches and normalized the scores. If the obtained score is greater than 0.5, we labeled the MRI as AD; otherwise, we labeled it as NC.

We have presented the results for both the ADNI and GARD datasets. The proposed model performed better on the GARD dataset in comparison to the ADNI dataset.

GARD dataset was collected only from Korean patients. Therefore, the brain structure and other anatomical factors are rationally homogeneous in each MRI. Any deviation and atrophy is comparatively easier to detect. In addition to this, the dataset only provides baseline MRIs, and the number of patients in the GARD dataset (326) is much higher than that in the ADNI dataset (we used only 60).

On the other hand, ADNI participants were recruited at 57 sites in the USA and Canada, with ages from 55 to 90. As the selection was from diverse races, ethnicities and age groups, the dataset includes heterogeneous brains. In addition to these characteristics, there are several scans for the same patient, where the progression from MCI to AD is also demonstrated. Therefore, these AD scans might impede the model's capability to obtain an accuracy equivalent to that of the GARD dataset. In the following subsections, we briefly outline the performance of each of the models.

A. HIPPOCAMPUS LOCATIONS

We manually navigated throughout the MRIs to observe the location of both hippocampi. We used the multi-image analysis GUI (Mango) [28] for this purpose. We selected six different (roughly equidistant) locations inside the hippocampus. We crosschecked the manual hippocampus locations, $H(h_x, h_y, h_z)$ by repeating the manual marking at different runs. We closely observed the detail views of three orthogonal planes to make sure that each patch contained the hippocampus. We considered a cube of side length 8, keeping each of the six different points as the center. From that cube, we semi-randomly drew locations to generate TVPs. We have also carefully avoided the repetition of locations. We show an example of the selected six points for an MRI in Figure 2. We show only the sagittal view here.

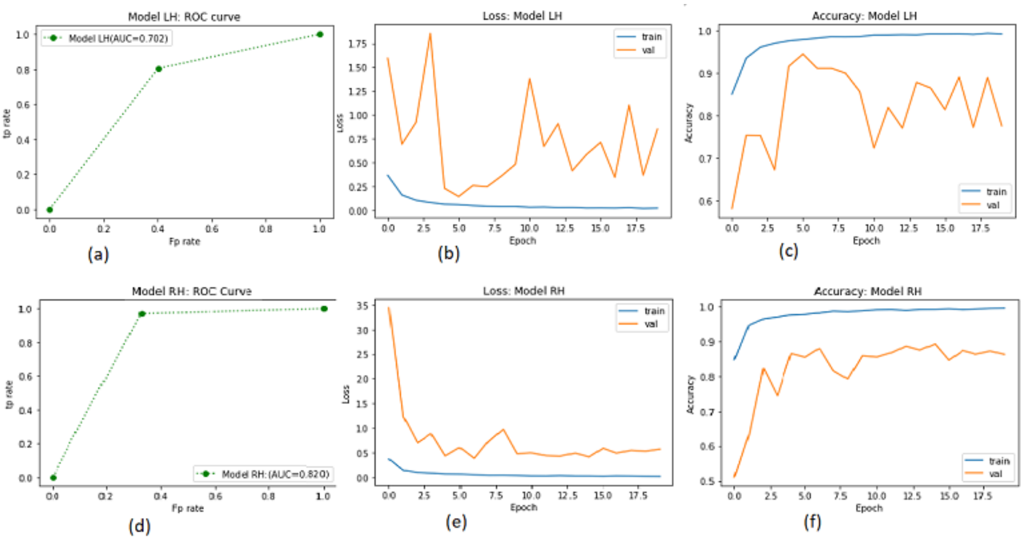

FIGURE 4. Examples of training and testing performance. (a) AUC scores of the test set for the left hippocampus (LH) model, (b) training and validation loss for the LH model, (c) training and validation accuracy for the LH model, (d) Area under receiver operating characteristics curve scores of the test set for the right hippocampus (RH) model, (e) training and validation loss for the RH model and (f) training and validation accuracy for the RH model.

B. RESULTS OF THE LEFT HIPPOCAMPUS CLASSIFIER

On the ADNI data, the presented model's overall accuracy for the left hippocampus classification was 80.40%. We tested the classifier on several runs with varying number of samples. During the prediction, we considered the class label for which the decision score is greater than 0.50. The reported one is found by feeding a balanced number of samples from both classes.

The area under the receiver operating characteristics curve for this model is 70.20%, as presented in figure 4. The model diagnosed 87.88% of the actual AD-affected MRIs as AD, and a total of 77.97% of the AD diagnosed MRIs are actually AD affected.

While testing with the GARD dataset, we obtained 83.27% overall accuracy. The precision, recall and f1 score for AD were measured as 76.88%, 90.50% and 84.34%, respectively; the measurements were 88.98%, 76.10% and 82.04%, respectively, for the NC class. The results indicate that 76.88% of the AD-classified MRIs are actually AD, and 90.50% of the actual AD-labeled MRIs are correctly classified.

C. RESULTS OF THE RIGHT HIPPOCAMPUS CLASSIFIER

The overall accuracy for classifying the right hippocampus as AD versus NC is 79.5% on the test set of the ADNI data. The number of samples drawn from the AD and NC class was kept balanced in the test set. Several test samples were generated for evaluating the model.

In total, 80% of the samples predicted as AD are true AD, and 79% of the AD-labeled samples are correctly classified by this model. The results are 78.89% and 79.69%, respectively, for the NC class. The assumption about the significance of the right hippocampus for AD diagnosis is strengthened by 82.0% area under the receiver operating characteristics curve (see figure 4(d)). The model demonstrated 81.56% accuracy on the GARD test data. The precision, recall and f1 score of the AD class were 76.88%, 87.56% and 81.87%, respectively; the scores were 87.09%, 76.12% and 81.24%, respectively, for the NC class.

D. RESULTS OF LEFT AND RIGHT HIPPOCAMPUS CLASSIFIER

We achieved 82.35% accuracy in classifying both hippocampi as AD versus NC. The precision, recall and f1 score are 83.30%, 82.37%, and 82.83% for the AD class, respectively, and 81.31%, 82.33%, and 81.82% for the NC class. The accuracy that we determined from the combined hippocampus classifier is greater than the individual hippocampus classifiers.

The performance for the GARD data also increased in the classification of both hippocampi. The overall accuracy was 86.28%. The details are shown in Table 2

E. RESULTS OF ENSEMBLES

The ensemble of the three models, as shown in Table 2, achieves 85.55% accuracy for the ADNI dataset. The ensemble provides precision, recall and the f1 score for the AD and NC class of approximately 85.5% ± 0.1%.

The overall accuracy for the GARD dataset is 90.05%. The precision for the AD class is 91.11%, while it is 88.85% for the NC class. In addition, 90.22% and 89.85 % are the recall scores for the AD and NC class, respectively. The f1 scores for the AD and NC class are 90.66% and 89.35%, respectively.

The findings demonstrated that the ensembles of the models provide better performance than the individual models.

TABLE 2. Results of the three TVP-based classifiers.

Dataset	Model	Class Label	Precision(%)	Recall(%)	f1 Score(%)	Accuracy(%)
ADNI	LH-Model	AD	77.97	87.15	92.30	80.40
		NC	83.81	72.98	78.02	
	RH-Model	AD	80.00	79.00	79.00	79.50
		NC	78.90	79.69	79.29	
	LHplusRHModel	AD	83.30	82.37	82.83	82.35
		NC	81.31	82.33	81.82	
	Ensemble Classifier	AD	85.66	85.52	85.59	85.55
		NC	85.43	85.57	85.50	
GARD	LH-Model	AD	76.88	90.50	84.34	83.27
		NC	88.98	76.10	82.04	
	RH-Model	AD	76.88	87.56	81.87	81.56
		NC	87.09	76.12	81.24	
	LHplusRHModel	AD	82.95	91.19	86.88	86.28
		NC	90.90	82.45	86.47	
	Ensemble Classifier	AD	91.11	90.22	90.66	90.05
		NC	88.85	89.85	89.35	

TABLE 3. Comparison of the proposed model with previous approaches.

Method	Modality	Sample Size	Whole Brain (Yes/No)	Accuracy(%
LPBoost+SVM [37]	MRI, FDG-PET	183,149	Yes	82
PCA, FDR+SVM [38]	MRI	345	Yes	76
SAE+MKSVM [2]	MR, PET	162	Yes	91±6
ResNet, VoxCNN [39]	MRI	111	Yes	79, 80
MMSDPN [6]	MRI, PET, CSF	103	Yes	97.13±4.44
PCA+RBM [7]	MRI, PET, CSF	103	Yes	91.4
SAE+3DCNN [24]	MRI	755	Yes	95.39
SAE+CNN [22]	MRI	432	Yes	93.80
Proposed Approach	MRI	352	No	85±5

As the decision scores of the individual models are added together and then softmax-normalized, each model's decision contributes to the final decision. The final decision-making procedure is a weighted voting strategy, as each decision score of an individual model can be considered as a weighted vote. This strategy alters the class label decided by individual classifiers with near to marginal scores. Therefore, the accuracy significantly increased after implementing an ensemble of the three models.

F. COMPARISON AND DISCUSSION

In previous methods, the whole brain and PET scan results were used to classify AD. The tendency is to use as much information as possible to train the model. In this paper, we show that we can obtain a comparable result using only the hippocampus. The comparative results for ADNI data are shown in Table 3.

In the multimodal stacked deep polynomial approach (MMSDPN) [6], as reported in the paper, the sparse connection in intermediate layers prevented overfitting. The authors used MRI, PET and cerebrospinal fluid (CSF) data to achieve 97.13% accuracy with 4.44% variation. The authors in [7] used the dropout technique to prevent overfitting in their multitask learning approach. Their approach demonstrated 91.4% accuracy on MRI, PET and CSF data. Both of the approaches were studied on the ADNI data (51 AD patients and 52 NC patients from each of the mentioned modalities).

The authors in [2] added a weight decay to regularize the objective function as a way to prevent overfitting. They experimented on 77 NC and 85 AD scans in both MRI and PET data from ADNI. Their model showed 82.59% accuracy with 5.33 variation in the MRI modality, which is 91.40%±5.56% in the multimodal data. Payan and Montana [23] studied the MRI modality with 755 scans for each of the classes. This approach also added a weight decay term similar to [2] to regularize the objective function in order to prevent overfitting. The method demonstrated 95.39% accuracy in classifying AD versus NC. A feature selection method was deployed along with an l-norm penalty on weights to prevent overfitting in [11]. The method achieved 82% accuracy with spatial augmentation on MRI and PET images. The reported accuracy without spatial augmentation on the data was 77%. This method was studied on 149 PET and 183 MRI images from

ADNI. To avoid the overfitting problem, Salvatore *et al.* [10] performed feature extraction and feature selection tasks separately for training-validation data and testing data. They reported 76% test accuracy on ADNI sMRI. The number of samples for AD was 137 and 162 for NC.

The method in [8] obtained 93.80% accuracy on ADNI MRI data. They did not explicitly discuss overfitting. This method learned a set of bases from natural images by deploying SAE and then used these bases to learn MRI features. The number of scans was 200 for AD and 232 for NC. VoxCNN in [35] reported 79.0% ± 0.08% accuracy. By using the residual neural network (ResNet), the approach [35] obtained 80.0% ± 0.07% accuracy. Here, the labeled sMRI scans included 50 AD and 61 NC from ADNI. Here, the overfitting problem was addressed by pre-training.

In this paper, the proposed patch-based ensembles of simple models demonstrate significant performance. We used only small patches (32 × 32) from the hippocampus of the brain MRIs and achieved comparable accuracy. Our patch generation reduces the scarcity of training data for generalization. Using the ensemble technique also contributed to building a robust model while avoiding the overfitting problem. It helps us to avoid obtaining an over-capacity network regarding the training time. The deployment of batch normalization [32] regularizes our models by enforcing that the inputs maintain a normal distribution in each layer. This regularization leads to better generalization. We also added dropout [36] in each model, which further address the overfitting problem.

To avoid problems of data imbalance, equal number of TVPs were generated from each of the majority(NC: 129 in ADNI and 171 in GARD) and minority (AD: 77 in ADNI and 81 in GARD) classes during training. We also avoided imbalance in the data during testing the individual models on TVP samples. For evaluating the model based on MRI, we kept the original ratio (20%) of the minority class dataset and took an equal number of MRIs from the majority class, i.e., we downsampled the majority class at test time.

VII. CONCLUSION

This work provides an efficient framework for AD diagnosis from brain MRI. We have considered the hippocampus, which is considered to be one of the most affected clinically studied biomarkers for AD detection. For the two different hippocampi in the brain, we had to deploy two patch-based classification models. However, deployment of another model for classifying both hippocampi increases the performance. We then designed ensemble models for an improved classification outcome. We designed the CNN classifiers based on TVPs on the semirandomly generated locations of the hippocampus region. This approach facilitated generation of the necessary data for training. After sufficient training, we combined the models to obtain the expected accuracy (85.55% for ADNI and 90.05% for GARD), which is comparable to the models designed in the MRI modality [1].

REFERENCES

[1] Alzheimer's Association, "2018 Alzheimer's disease facts and figures," *Alzheimer's Dementia*, vol. 14, no. 3, pp. 367–429, Mar. 2018.

[2] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, D. Feng, M. J. Fulham, and ADNI, "Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1132–1140, Apr. 2015.

[3] S. B. Hendrix, "Measuring clinical progression in MCI and pre-MCI populations: Enrichment and optimizing clinical outcomes over time," *Alzheimer's Res. Therapy.*, vol. 4, no. 4, p. 24, Jul. 2012.

[4] *World Alzheimer Report 2018*, Alzheimer's Diseases Int., London, U.K., Sep. 2018.

[5] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, Apr. 2011.

[6] J. Shi, X. Zheng, Y. Li, Q. Zhang, and S. Ying, "Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 173–183, Jan. 2018.

[7] F. Li, L. Tran, K. H. Thung, S. Ji, D. Shen, and J. Li, "A robust deep model for improved classification of AD/MCI patients," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 5, pp. 1610–1616, Sep. 2015.

[8] A. Gupta, M. Ayhan, and A. Maida, "Natural image bases to represent neuroimaging data," in *Proc. 30th Int. Conf. Int. Conf. Mach. Learn.*, Atlanta, GA, USA, Jun. 2013, pp. 1–8. [Online]. Available: http://jmlr.org/proceedings/papers/v28/gupta13b.html

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015. doi: 10.1038/nature14539.

[10] C. Salvatore, A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, and I. Castiglioni, "Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach," *Frontiers Neurosci.*, vol. 9, p. 307, Sep. 2015.

[11] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M. K. Chung, and S. C. Johnson, "Spatially augmented LPboosting for AD classification with evaluations on the ADNI dataset," *NeuroImage*, vol. 48, no. 1, pp. 138–149, 2009.

[12] T. K. Khan, "Introduction to Alzheimer's disease biomarkers," in *Biomarkers in Alzheimer's Disease*, 1st ed. New York, NY, USA: Academic, 2016, p. 13.

[13] D. Chan, N. C. Fox, R. I. Scahill, W. R. Crum, J. L. Whitwell, G. Leschziner, A. M. Rossor, J. M. Stevens, L. Cipolotti, and M. N. Rossor, "Patterns of temporal lobe atrophy in semantic dementia and Alzheimer's disease," *Ann. Neurol.*, vol. 49, no. 4, pp. 433–442, Apr. 2001.

[14] C. R. Jack, Jr., R. C. Petersen, Y. Xu, P. C. O'Brien, G. E. Smith, R. J. Ivnik, E. G. Tangalos, and E. Kokmen, "Rate of medial temporal lobe atrophy in typical aging and Alzheimer's disease," *Neurology*, vol. 51, no. 4, pp. 993–999, Oct. 1986.

[15] D. Shen, C.-Y. Wee, D. Zhang, L. Zhou, and P.-T. Yap, "Machine learning techniques for AD/MCI diagnosis and prognosis," in *Machine Learning in Healthcare Informatics*. Berlin, Germany: Springer, 2014.

[16] Y. Wang, M. Liu, L. Guo, and D. Shen, "Kernel-based multi-task joint sparse classification for Alzheimer's disease," in *Proc. IEEE 10th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2013, pp. 1364–1367.

[17] S.-T. Yang, J.-D. Lee, T.-C. Chang, C.-H. Huang, J.-J. Wang, W.-C. Hsu, H.-L. Chan, Y.-Y. Wai, and K.-Y. Li, "Discrimination between Alzheimer's disease and mild cognitive impairment using SOM and PSO-SVM," *Comput. Math. Methods Med.*, vol. 2013, Apr. 2013, Art. no. 253670.

[18] K. R. Gray, P. Aljabar, R. A. Heckemann, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage*, vol. 65, pp. 167–175, Jan. 2013.

[19] A. Ortiz, J. M. Górriz, J. Ramírez, and F. J. Martínez-Murcia, and ADNI, "LVQ-SVM based CAD tool applied to structural MRI for the diagnosis of the Alzheimer's disease," *Pattern Recognit. Lett.*, vol. 34, pp. 1725–1733, Oct. 2013.

[20] J. Escudero, J. P. Zajicek, E. Ifeachor, and ADNI, "Machine Learning classification of MRI features of Alzheimer's disease and mild cognitive impairment subjects to reduce the sample size in clinical trials," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, Aug./Sep. 2011, pp. 7957–7960.

[21] R. K. Lama, J. Gwak, J.-S. Park and S.-W. Lee, "Diagnosis of Alzheimer's disease based on structural MRI images using a regularized extreme learning machine and PCA features," *J. Healthcare Eng.*, vol. 2017, Jun. 2017, Art. no. 5485080. doi: 10.1155/2017/5485080.

[22] T. Brosch and R. C. Tam, "Manifold learning of brain MRIs by deep learning," in *Proc. 16th Int. Conf. Med. Image Comput. Comput.-Assisted Intervent. (MICCAI)*, Nagoya, Japan, Sep. 2013, pp. 633–640.

[23] A. Payan and G. Montana, "Predicting Alzheimer's disease: A neuroimaging study with 3D convolutional neural networks," in *Proc. Int. Conf. Pattern Recognit. Appl. Methods,* Lisbon, Portugal, 2015, pp. 355–362.

[24] N. Srivastava, G. Hinton, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[25] M. Cercignani, N. G. Dowell, and P. S. Tofts, *Quantitative MRI of the Brain: Principles of Physical Measurement*. New York, NY, USA: CRC Press, 2018.

[26] G. Rudow, O. Pletnikova, B. Crain, J. C. Troncoso, D. Iacono, R. O'Brien, A. B. Zonderman, S. M. Resnick, and Y. An, "Mild cognitive impairment and asymptomatic Alzheimer disease subjects: Equivalent β-amyloid and tau loads with divergent cognitive outcomes," *J. Neuropathol. Exp. Neurol.*, vol. 73, no. 4, pp. 295–304, 2014.

[27] R. A. Sperling *et al.*, "Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," *Alzheimer's Dementia*, vol. 7, no. 3, pp. 280–292, 2011.

[28] University of Texas Health Science Center. *Mango 2006–2019*. Accessed: Jun. 6, 2019. [Online]. Available: ric.uthscsa.edu/mango/

[29] L. Deng and Y. Dongg, "Deep learning: Methods and applications," in *Plastics*, J. Peters, Ed. Hanover, MA, USA: Now, 2014, pp. 02–06.

[30] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, Haifa, Israel, Jun. 2010, pp. 807–814.

[31] J. Nagi, F. Ducatelle, G. A. Di Caro, D. C. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Kuala Lumpur, Malaysia, Nov. 2011, pp. 342–347.

[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, Lille, France, Jul. 2015, pp. 448–456.

[33] P. D. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2016, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[34] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist. (AISTATS)*. Sardinia, Italy: Chia Laguna Resort, May 2010, pp. 249–256.

[35] S. Korolev, A. Safiullin, M. Belyaev, Y. Dodonova, A. Quattrone, and I. Castiglioni, "Residual and plain convolutional neural networks for 3D brain MRI classification," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (IEEE-ISBI)*, Apr. 2017, pp. 835–838.

[36] P. Baldi and P. J. Sadowski, "Understanding dropout," in *Proc. Adv. Neural Inf. Process. Syst., 27th Annu. Conf. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, Dec. 2013, pp. 2814–2822.

[37] A. Kessy, A. Lewin, and K. Strimmer, "Optimal whitening and decorrelation," Dec. 2016, *arXiv:1512.00809*. [Online]. Available: https://arxiv.org/abs/1512.00809

[38] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.

SAMSUDDIN AHMED received the B.Sc. degree in computer science and engineering from the University of Chittagong, in 2010. He is currently pursuing master degree with the Computer Vision Laboratory, Department of Computer Engineering, Chosun University, South Korea. In 2010, he joined as a Lecturer in CSE with the State University of Bangladesh. In 2011, he joined as a Lecturer in CSE with the Bangladesh University of Business and Technology. In 2011, he was promoted as an Assistant Professor. His research interests include machine vision, deep learning, data analysis, securities, and AI.

KYU YEONG CHOI received the Ph.D. degree in life science from GIST, Gwangju, South Korea, in 2003. He is currently a Research Professor with the National Research Center for Dementia, Chosun University, Gwangju. His research interests include early detection of Alzheimer's disease (AD) and the role of genetic variants on the onset of AD.

JANG JAE LEE received the Ph.D. degree from the Department of Computer Science and Statistics, Chosun University, Gwangju, South Korea, in 2007, where he is currently a Research Professor with the National Research Center for Dementia. His current research interests include Genome-wide association studies and imaging genetics for Alzheimer's disease.

BYEONG C. KIM received the medical doctor degree, and the M.S. and Ph.D. degrees from the Department of Neurology, Chonnam National University Medical School, Gwangju, South Korea, 1991, 1994, and 2000, respectively, where he is currently a Professor. His current research interests include neuroimage (structural and molecular) and CSF biomarkers in patients with neurodegenerative diseases.

GOO-RAK KWON received the M.S. degree from the School of Electrical and Computer Engineering, SungKyunKwan University, in 1999, and the Ph.D. degree from the Department of Mechatronic Engineering, Korea University, in 2007. He has served as Chief Executive Officer and the Director of Dalitech Co. Ltd., from 2004 to 2007. He joined the Department of Electronic Engineering, Korea University, where he was a Postdoctoral Researcher supporting the BK21 Information Technique Business, from 2007 to 2008. He has been a Professor with Chosun University, since 2017. He has also been an Associate Dean with the Industry-academic Cooperation Foundation, since 2018. He has contributed 44 and 75 articles to journals and conference proceedings, respectively. He also holds 27 patents on the medical image analysis and the security of multimedia contents for digital rights management. He was a member of the IEICE, and IS&T in the international institute. In the domestic institute, he was a member of the signal processing society in the IEIE, KMMS, KIPS, and KICS. His research interests include medical image analysis, A/V signal processing, video communication, and applications.

KUN HO LEE received the B.S. degree from the Department of Genetic Engineering, Korea University, Seoul, South Korea, in 1989, the M.S. and Ph.D. degrees from the Department of Molecular Biology, Seoul National University, Seoul, in 1994 and 1998, respectively. He is currently an Associate Professor with the Department of Biomedical Science, Chosun University, Gwangju, South Korea. He is also with the National Research Center for Dementia, Chosun University. His current research interests include the brain image analysis, and the development of prediction model for neurodegenerative diseases based on MR imaging and genetic variants.

HO YUB JUNG received the B.S. degree in electrical engineering from The University of Texas at Austin, in 2002, and the M.S. and Ph.D. degrees in electrical engineering and computer science from Seoul National University, in 2006 and 2012, respectively. He was with Samsung Electronics for two years as a Senior Engineer. Since 2017, he has been an Assistant Professor with the Department of Computer Engineering, Chosun University. His research interests include computer vision, machine learning, and medical imaging.

• • •